AD 429802

64-9

BIOMETRICS UNIT
DEPARTMENT OF PLANT BREEDING

NEW YORK STATE COLLEGE OF AGRICULTURE

CORNELL UNIVERSITY

ITHACA, NEW YORK

NOTE ON THE PROPORTION OF GENETIC DEVIATES IN THE TAILS OF A NORMAL POPULATION

Technical Report No. 13

Department of Navy

Office of Naval Research

Contract No. Nonr-401(39)
Project No. (NR 042-212)

D. S. Robson
Biometrics Unit
N. Y. S. College of Agriculture
Cornell University
Ithaca, New York

LeRoy Powers
Sugarbeet Investigations, U.S.D.A.
Agricultural Research Service
University of Colorado
Fort Collins, Colorado

This work was supported in part by the Office of Naval Research.

NOTE ON THE PROPORTION OF GENETIC DEVIATES IN THE TAILS OF A NORMAL POPULATION

D. S. Robson [1] and LeRoy Powers [2]

Introduction

The phenotypic array exhibited by a segregating genetic population reflects both the genetic and the environmental variability within the population. As a consequence, an element of uncertainty attaches to selection for genetically superior individuals on the basis of their phenotypic traits. The latter may, by chance, be merely the result of an unusually favorable environment acting upon a genotype which under less favorable conditions would display only a mediocre or even undesirable phenotype. Chances for the occurence of such phenotypic deception depend, of course, upon the magnitude of the environmentally induced variability as compared to that due to genetic differences.

Any mathematical formulation of this problem to allow the geneticist to numerically evaluate his chances for successful selection requires a detailed description of the phenotypic frequency distribution in the population. Structurally, the total segregating population may be regarded as a mixture of subpopulations, with each subpopulation representing the distribution of phenotypes produced by a single genotype under the existing range of environmental conditions, and with each subpopulation or genotype contributing to the total population in proportion to its genotypic frequency. A mathematical description of the population therefore consists of specifying the relative frequency of each genotype and the exact form of its associated distribution of phenotypes.

Under most circumstances where selection is practiced for economic purposes a large number of both genetic and environmental factors operate

[1] Associate Professor of Biological Statistics, Cornell University.

[2] Geneticist, Crops Research Division, Agricultural Research Service, U. S. Department of Agriculture.

at variable levels to determine the phenotypes appearing in the population. Empirical evidence supports the belief that in this case the total frequency distribution and also the component distributions for a quantitative phenotypic trait are approximately Gaussian in form. A standard population model which has therefore come into use in such problems as the prediction of advancement under selection is Eisenhart's Model II (1947) representing, in the simplest case, a normal mixture of normal subpopulations with constant variance. Each genotype is assumed to generate a normal distribution of phenotypes under the existing range of environmental conditions, and the distribution of phenotypic means (called genotypic values) is likewise normal.

Graphs of the Proportion of Genetic Deviates

The phenotypic value X for some quantitative trait of an individual selected at random from a genetic population may be regarded conceptually as the sum of two components,

G = average phenotype for the genotype of the chosen individual

E = deviation of the particular phenotype of the chosen individual from the average phenotype (G) for the genotype of that individual = X-G

or

$$X = G + E$$

The first component G is conventionally called the genotypic value and E is the environmental effect. If the population structure is a normal mixture of normal subpopulations having a common environmental variance then the chance variables G and E follow independent normal distributions, G having a mean value of $\bar{g}$ and variance σ_g^2, E having a mean value of zero and variance σ_e^2, so that X itself follows a normal distribution with mean $\bar{g}$ and variance $\sigma_g^2 + \sigma_e^2$.

A probability distribution of particular interest to the geneticist is the a posteriori distribution of G among individuals of a fixed phenotype x; that is, given that he has selected an individual of phenotype x, the geneticist is then concerned with the probability that this chosen

individual is of superior genotype, say greater than some standard value g. For any selected phenotypic value x the conditional distribution of genotypic values G is in this case normal with mean $\bar{g} + h\,(x-\bar{g})$ and variance $\sigma_g^2\ (1-h)$, where h is the heritability ratio

$$h = \frac{\sigma_g^2}{\sigma_e^2 + \sigma_g^2}$$

or the ratio of genetic to total variance in the population.
The desired probability that the genotypic value will exceed some specified value g is therefore given by the standard cumulative normal probability $\Phi\,([\bar{g}+h(x-\bar{g})-g] / \sigma_g\sqrt{1-h})$ or, expressing x and g in standard units as

$$x' = \frac{x-\bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}} \quad \text{and} \quad g' = \frac{g-\bar{g}}{\sigma_g}$$

we obtain the simplified form

$$\Phi([x'\sqrt{h}-g'] / \sqrt{1-h}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x'\sqrt{h}-g}{\sqrt{1-h}}} e^{\frac{t^2}{2}}\,dt$$

which depends on the variance components only through the heritability ratio. A special case of some interest is where $g = \bar{g}$, giving $\Phi(x' / \sqrt{\frac{1-h}{h}})$ as the probability that an individual of phenotype x will be genetically above average; note that when $h = \frac{1}{2}$ this reduces to

$\Phi(x')$ = area to the left of x' under the standard normal curve.

For example, with $x' = 1.645$ and $h = \frac{1}{2}$, the probability that the genotypic value of the selected individual exceeds the population mean is $\Phi(1.645) = .95$; if h were $\frac{1}{4}$ instead of $\frac{1}{2}$ then the desired probability would be $\Phi(1.645/\sqrt{3}) = \Phi(.950) = .83$, while $h = 3/4$ would give $\Phi(1.645\sqrt{3}) = \Phi(2.849) = .998$. These answers may be interpolated from Figure 1 where the general form of the solution is illustrated by plotting $\Phi([x'\sqrt{h}-g'] / \sqrt{1-h})$ as a function of g' with x' fixed at the upper 5 percent tail value of 1.645 and with h taking the values 0.1 to 0.9 by steps of 0.1. Figures 2 and 3 display the corresponding curves with x' fixed at the upper 10 percent value of 1.282 and the upper quartile value of .674, respectively.

Where selection operates on the entire upper tail of the phenotypic distribution, taking all individuals whose phenotypic value exceeds x, the question of interest becomes the proportion of these selected individuals having genotypic values exceeding g. More precisely, this proportion may be interpreted as the conditional probability that a randomly selected individual exhibiting a phenotypic value greater than x will also have a genotypic value greater than g. Again, the solution to this problem is expressible in integral form as

$$P_{g \cdot x'} = \frac{1}{1-\Phi(x')} \int_{x'}^{\infty} \Phi\left(\frac{y\sqrt{h}-g'}{\sqrt{1-h}}\right) d\Phi(y)$$

though in this case the integral is not previously tabulated. The only non-trivial case which permits further analytic reduction is $g' = 0$ and $h = \frac{1}{2}$; when half of the total varience is genetic then the probability that an individual randomly selected from the region $X > x$ will be genotypically above average is $[1 + \Phi(x')]/2$. Thus, in this case, the solution may be read directly from the cumulative normal tables; for example, the proportion of above average genotypes in the upper 5 percent tail of the phenotypic distribution is then

$$[1 + \Phi(1.645)]/2 = [1 + .95]/2 = .975.$$

A numerical solution in the general case can be obtained only by numerical integration, for which the most convenient approach is to hold g' and h fixed and allow x' to vary. Graphs of this probability function of g' appearing in Figures 4-12 are therefore plotted as functions of x', with each curve corresponding to a fixed value of g' and h. For example, from the curve for $P_{g'} = .05$ ($g' = 1.645$) and $h = \frac{1}{2}$ we see that in the upper 25 percent tail of the phenotypic distribution ($x' = .674$) the proportion of standardized genetic deviates exceeding $g' = 1.645$ is .17; in other words, 17 percent of the phenotypic top 25 percent also belong to the genotypic top 5 percent of the population when half of the variability is genetic.

A more complete picture of the expected proportions of genetic deviates may be constructed in histogram form by a slight extension of the results in Figures 4-12. For a set of phenotypic intervals

$(x_1, x_2), \ldots, (x_g, +\infty)$ on the standard scale

$$x_i = \frac{X_i - \bar{g}}{\sqrt{\sigma g^2 + \sigma e^2}}$$

the proportion $P_g(x_i, x_{i+1})$ of genotypic values exceeding the population mean by an amount $g\sigma_g$ can be computed from the formula

$$P_g(x_i, x_{i+1}) = \frac{P_{g \cdot x_i}\,[1-\bar{\Phi}(x_i)] - P_{g \cdot x_{i+1}}\,[1-\bar{\Phi}(x_{i+1})]}{[1-\bar{\Phi}(x_i)] - [1-\bar{\Phi}(x_{i+1})]}$$

Again referring to Figure 8 for the case $h = 0.5$, we illustrate this procedure with intervals of length $x_{i+1} - x_i = 0.5$ between $x_1 = .25$ and $x_8 = x_9 = +2.25$. Taking the interval (1.25, 1.75), for example, we find (either by visual interpolation on the right hand scale of Figure 8 or directly from tables of the standard cumulative normal distribution)

$$1-\bar{\Phi}(1.25) = .1056 \qquad 1-\bar{\Phi}(1.75) = .0401$$
$$[1-\bar{\Phi}(1.25)] - [1-\bar{\Phi}(1.75)] = .0655$$

Then to find the expected proportion of above average genotypes in this interval we refer to the $P = 0.50$ (or $g = 0$) curve in Figure 8; at $x_i = 1.25$ this gives

$$P_{g \cdot x_i} = P_{0 \cdot 1.25} = .947$$

and at $x_{i+1} = 1.75$,

$$P_{g \cdot x_{i+1}} = P_{0 \cdot 1.75} = .980$$

so that

$$P_g(x_i, x_{i+1}) = P_0(1.25, 1.75) = \frac{.947(.1056) - .980(.0401)}{.0655}$$

$$= \frac{.0607}{.0655}$$

$$= .9267$$

Thus, when $h = \frac{1}{2}$, 6.55 percent of the population falls in the phenotypic interval

$$1.25 < \frac{X - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}} < 1.75$$

and 92.67 percent of the individuals falling in this interval are genotypically above average.

Similarly, from the curve for P = 0.25 (g = .6745) in Figure 8 we find

$$P_{g \cdot x_i} = P_{.6745 \cdot 1.25} = .756 \qquad P_{g \cdot x_{i+1}} = P_{.6745 \cdot 1.75} = .867$$

so that

$$P_g(x_i, x_{i+1}) = P_{.6745}(1.25, 1.75) = \frac{.756(.1056) - .867(.0401)}{.0655}$$

$$= \frac{.0451}{.0655}$$

$$= .6885$$

Thus, 68.85 percent of the individuals in this interval belong to the top quartile of the genotypic distribution. By subtraction, 92.67 - 68.85 = 23.82 percent of the individuals in this interval belong to the second quartile of the genotypic distribution. The remaining computations for this and other intervals were performed in the same manner and are shown in Table 1. The resulting histogram is plotted in Figure 13, indicating for each phenotypic class the proportions of individuals belonging to the various percentiles of the genotypic distribution.

Table 1 Percentage of individuals in a phenotypic class which belong to the top P_g percent of the genotypic distribution when $h = \frac{1}{2}$.

Phenotypic class interval	Phenotypic frequency (percent)	Percentage of the class frequency belonging to the top P_g percent of the genotypic distribution				
		$P_g = 50$	$P_g = 25$	$P_g = 10$	$P_g = 5$	$P_g = 1$
.25 to .75	17.47	68.63	31.54	9.56	3.43	0.29
.75 to 1.25	12.10	83.39	50.17	20.58	9.09	1.07
1.25 to 1.75	6.55	92.67	68.85	36.79	19.85	3.51
1.75 to 2.25	2.79	97.49	83.51	55.91	35.84	9.32
2.25 to ∞	1.22	99.18	94.26	77.05	59.84	25.41

The Effect of Population Size: Before Selection

The preceding results describe some characteristics of an abstract infinite population, while the genetic population actually observed and selected from is of finite size N, representing only a sample from this potential infinite population. Finiteness of the observed population has no effect on the results plotted in Figures 1 - 3, which are conditional on a single selected phenotypic value; but the application of Figures 4 - 12, which are conditional on a selected upper tail of the phenotypic distribution, requires further explanation in the finite case.

A selection procedure which takes all individuals in the tail of a distribution may be defined in essentially two different ways, by specifying either (i) the minimum acceptable phenotypic value or (ii) the percentile of the observed phenotypic distribution at which selection begins. When the minimum acceptable phenotype is fixed in advance then the number or percentage of the prospective population of size N which will be selected is a chance variable, unknown in advance of the selection experiment, while if the selection rate or percentage is fixed then the minimum phenotypic value which will be accepted is a chance variable and unknown in advance of the experiment. In either case, chance variations in the respective unknown quantities will decrease as N is increased, and as N approaches infinity the two procedures become equivalent; that is, when population size is infinite, the specification that selection will take all individuals with phenotypic values exceeding x is equivalent to the specification that selection will take a fraction $1-\Phi(x)$ from the upper tail of the phenotypic distribution. This asymptotic equivalence is expressed in Figures 4 - 12 by labeling the ordinate with both the scale of x and $1-\Phi(x)$.

For the purposes of planning a finite selection experiment of either type, the potential magnitude of chance fluctuations in the respective unknown quantities must be considered. For example, if a type (i) experiment is contemplated with the minimum acceptable phenotype at some preassigned level x then rational planning requires that the population size N be chosen large enough to provide reasonable assurance that at least one of the N phenotypes will exceed x. The preassigned value of x

in this case is presumably based on considerations of the facts revealed in Figures 1 - 12, and is likewise chosen to provide reasonable assurance that an individual of this phenotype will be of superior genotype.

Analysis of the type (i) experiment model shows that the probability of finding at least one phenotype exceeding $x = (X-\bar{g}) / \sqrt{\sigma_g^2 + \sigma_e^2}$ in a population of size N is $1 - [\Phi(x)]^N$, and the probability of finding one whose genotype also exceeds some specified value $g = (G-\bar{g})/\sigma_g$ is $1 - [1-P_{g \cdot x}(1-\Phi(x))]^N$. Thus, in order to obtain $100(1-\alpha)$ percent assurance that the type (i) experiment will produce at least one selection, population size N must be chosen to satisfy the equation

$$1 - [\Phi(x)]^N = 1 - \alpha$$

or

$$N = \frac{\log \alpha}{\log \Phi(x)}$$

In order to provide the same assurance of obtaining at least one selection with a genotypic value exceeding any specified value g, a larger value of N is required,

$$N = \frac{\log \alpha}{\text{Log}[1-P_{g \cdot x}(1-\Phi(x))]}$$

For example, if all phenotypes greater than two standard deviations above the mean are to be selected ($x = 2, \Phi(x) = .9773$) then in order to provide 90 percent assurance ($\alpha = 0.1$) of obtaining at least one selection from the top 5 percent of the genotypic distribution ($P_g = .05$, $g = 1.645$) the population size N must be chosen as follows

h	$P_{g \cdot x}$	N
0.1	.1744	580
0.2	.2597	390
0.3	.3427	295
0.4	.4270	235
0.5	.5146	196
0.6	.6071	165
0.7	.7062	143
0.8	.8139	124
0.9	.9293	108
1.0	1	101

The values of $P_{\xi \cdot x}$ in these computations were those obtained by numerical integration, but may be read with two-digit accuracy from Figures 4 - 12, respectively. As an illustration, referring to Figure 8 with h = 0.5 we find that the curve for $P_\xi = .05$ intersects with x = 2 at $P_{\xi \cdot x} \doteq .515$, giving

$$N = \frac{\log_{10} 0.1}{\log_{10} [1-.0227(.515)]} = \frac{-1}{\log_{10} 0.9883} = \frac{1}{.00511} = 196$$

A more detailed characterization of a contemplated type (i) experiment is given by the probability of selecting exactly m individuals belonging to the top 100 P_ξ percent of the genotypic distribution, which is the binomial probability

$$\binom{N}{m} P_{\xi \cdot x}^{m} [1-\Phi(x)]^{m} \; [1-P_{\xi \cdot x}(1-\Phi(x))]^{N-m}$$

or approximately

$$\frac{[NP_{\xi \cdot x}(1-\Phi(x)]^{m}}{m!} e^{-NP_{\xi \cdot x}(1-\Phi(x))}$$

Thus, for the case illustrated above, with N = 196 the probability of obtaining exactly one selection from the genetic top 5 percent is

$$\binom{196}{1}(.0227)(.515)(.9883)^{195} = 2.291338\ (.1008) = .231$$

or by the approximation

$$2.291338\ (.1011) = .232$$

Similarly, the chance of obtaining exactly two selections of this kind is approximately

$$\frac{(2.291338)^2}{2}\ (.1011) = .265$$

while for m = 3

$$\frac{(2.291338)^3}{3(2)}\ (.1011) = .203$$

and for m = 4

$$\frac{(2.291338)^4}{4(3)(2)}\ (.1011) = .116$$

and so on, the probabilities from m = 1 onward adding to the previously specified $1-\alpha = .90$. The expected number of such selections is $NP_{g\cdot x}\ (1-\Phi(x))$, or in this case 2.291338.

If a type (ii) experiment is contemplated with a fixed selection rate of 100 $(1-\Phi)$ percent then the population size N should be chosen large enough to provide reasonable assurance $(1-\alpha)$ that the $k = N(1-\Phi)$ selections are genetically superior, or at least that the best of these selected phenotypes is a genetically superior individual. A characterization of the type (ii) model by means of the probability distribution of m, the number of selections exceeding g in genotypic value, is readily accomplished analytically, but the form of the distribution is quite cumbersome computationally. A more convenient characterization is given by the genotypic distribution associated with the minimum or with the maximum selected phenotype, the latter being equivalent to the distribution of m for the special case $k = N(1-\Phi) = 1$.

The genotypic distribution associated with the k'th ranking phenotype in a population of size N is

$$1-P_{g,k} = \int_{-\infty}^{\infty} \Phi\left(\frac{g-x\sqrt{h}}{\sqrt{1-h}}\right)\ d\sum_{r=0}^{k-1}\binom{N}{r}\ [1-\Phi(x)]^r\ [\Phi(x)]^{N-r}$$

where $P_{g,k}$ denotes the probability that the genotype of the phenotypically k'th largest individual will exceed the population mean by at least an amount $g\sigma_g$. For k = 1, or for the largest of N phenotypes, this becomes

$$P_{g,1} = \int_{-\infty}^{\infty} \Phi\left(\frac{x\sqrt{h}-g}{\sqrt{1-h}}\right)\, d\Phi^N(x)$$

Numerical evaluation of the integral $P_{g,k}$ for the purpose of appraising a contemplated type (ii) experimental plan is still somewhat tedious, but arbitrarily close bounds on the integral may be computed as

$$\sum_{i=0}^{n} \Phi\left(\frac{x_i-g\sqrt{h}}{\sqrt{1-h}}\right)[F_k(x_{i+1})-F_k(x_i)] < P_{g,k} < \sum_{i=0}^{n} \Phi\left(\frac{x_{i+1}-g\sqrt{h}}{\sqrt{1-h}}\right)[F_k(x_{i+1})-F_k(x_i)]$$

where $-\infty = x_0 < x_1 < \cdots < x_n < x_{n+1} = +\infty$ is any judiciously chosen montone sequence and

$$F_k(x) = \sum_{r=0}^{k-1} \binom{N}{r} [1-\bar{\alpha}(x)]^r \, [\bar{\alpha}(x)]^{N-r}$$

Furthermore, when k is small and N is large, $F_k(x)$ remains extremely small until $\bar{\alpha}(x)$ gets near unity, and then $F_k(x)$ is closely approximated by

$$F_k(x) \doteq e^{-N[1-\bar{\alpha}(x)]} \sum_{r=0}^{k-1} \frac{[N(1-\bar{\alpha}(x)]^r}{r!}$$

In particular, when k = 1,

$$F_1(x) = \bar{\alpha}^N(x) \doteq e^{-N[1-\bar{\alpha}(x)]}$$

remains less than .005 until $\bar{\alpha}(x)$ attains the value

$$\bar{\alpha}(x_1) = 1 - \frac{5.29832}{N}$$

and increases to .995 at

$$\bar{\alpha}(x_n) = 1 - \frac{.00501}{N}$$

thus indicating a judicious range for the sequence $x_1 < \cdots < x_n$.

Certain special cases do exist where the integral $P_{g,k}$ can be evaluated explicitly, the most interesting being the case $h = \frac{1}{2}$ and $g = 0$. When half of the phenotypic variability in the (infinite) population is of genetic origin then the probability that the k'th ranking phenotype in a population of size N will be genetically above average is

$$P_{0,k} = 1 - \frac{k}{N+1}$$

From this result it follows that the expected number of above average genotypes among the best k out of N phenotypes is

$$\sum_{j=1}^{k} P_{0,j} = k\left[1 - \frac{k+1}{2(N+1)}\right]$$

A second but rather trivial special case arises when h = 1, giving

$$P_{g,k} = 1 - F_k(g)$$

which is readily evaluated from tables of the standard normal distribution. While this limiting case is of no particular genetic interest it does

provide a convenient bound which may be useful as a check against numerical integration at other values of h.

Numerical integration employing the devices mentioned above was carried out for the case k = 1 to obtain a solution to $P_{s,1} = 0.9$ as an equation in N. These results, presented in Table 1 as a guide to experiment planning, indicate the population size required for 90 percent confidence that the first ranking selection will belong to the top 5, 10, 25 or 50 percent of the genetic population.

Table 2 Population size required for 90 percent certainty that the genotypic value of the phenotypically best individual will fall in a specified upper percentile of the genotypic distribution.

Heritability $h = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$	Percentile of the genotypic value of the best phenotype			
	Top 50%	Top 25%	Top 10%	Top 5%
0.1	12,137	7×10^8	2×10^{14}	10^{18}
0.2	147	31,140	3×10^7	5×10^9
0.3	35	1,070	84,145	2×10^8
0.4	16	193	4,650	5×10^4
0.5	9	80	870	5,300
0.6	7	34	254	1,060
0.7	6	17	109	415
0.8	5	14	58	165
0.9	4	11	34	82
1.0	4	8	22	45

The Effect of Population Size: After Selection

At the planning stage of a selection experiment the phenotypes to be selected are yet unknown, except in the form of a lower bound in the case of a type (i) experiment, and upon completion of the phenotypic selection the genotypic values of the chosen individuals are yet unknown. Figures 1-12 are therefore still of some interest to the geneticist at the post-selection stage and, in particular, Figures 1-3 apply to describe the genotypic

distribution associated with any selected phenotypic value, regardless of the rank of that individual among the selections. If phenotypic values are not actually measured in the selection experiment but only compared to a standard value x in a type (i) experiment (i.e. phenotypes are observed only to be greater or less than x) or simply ranked in a type (ii) experiment then Figures 1-3 do not apply, and the use of Figures 4-12 depends more specifically on the type of information available on the selected phenotypes.

In a type (i) experiment where no information is obtained other than the exact number k of phenotypes exceeding the standard value x then for a random one of these k individuals the computations of Figures 4-12 apply directly, and for all k individuals a binomial distribution applies. The probability that m of the selected k individuals fall in the top 100 P_g percent of the genotypic distribution is simply

$$\binom{k}{m} P_{g\cdot x}^{m}(1-P_{g\cdot x})^{k-m}$$

For example, if k = 10 individuals are found to be phenotypically larger than a preassigned standard value of x = 1.645 in a population with heritability h = 0.4 then the probability that at least 3 of these 10 belong to the top 100 P_g = 5 percent of the genotypic distribution is

$$1-\binom{10}{0}P_{g\cdot x}^{0}(1-P_{g\cdot x})^{10}-\binom{10}{1}P_{g\cdot x}^{1}(1-P_{g\cdot x})^{9}-\binom{10}{2}P_{g\cdot x}^{2}(1-P_{g\cdot x})^{8}$$

From Figure 7 and the curve P_g = 5 at x = 1.645 we find $P_{g\cdot x}$ = .333, giving

$$1 - .667^{10} - 10(.333)(.667)^{9} - (.333)^{2}(.667)^{8} = .6996$$

Notice that knowledge of k will, on the average, decrease the variance of m by a factor of $(1-P_{g\cdot x})/[1-P_{g\cdot x}+P_{g\cdot x}\bar{\Phi}(x)]$, or approximately by $1-P_{g\cdot x}$.

Additional information concerning the ordering among the k selected phenotypes in a type (i) experiment is difficult to analyze numerically because of integration problems. The genotypic distribution associated with the lowest ranking selected phenotype, for example, is given by the integral

$$\frac{k}{[1-\bar{\Phi}(x)]^{k}}\int_{x}^{\infty}\bar{\Phi}\left(\frac{g-y\sqrt{h}}{\sqrt{1-h}}\right)[1-\bar{\Phi}(y)]^{k-1}d\bar{\Phi}(y)$$

while that for the highest ranking phenotype is

$$\frac{k}{[1-\bar{\Phi}(x)]}\, k \int_{x}^{\infty} \bar{\Phi}\left(\frac{g-y\sqrt{h}}{\sqrt{1-h}}\right) [\bar{\Phi}(y)-\bar{\Phi}(x)]^{k-1} d\bar{\Phi}(y)$$

Such functions could be integrated numerically by the methods employed in computing Table 1; however, the number of cases to be considered is quite large, requiring extensive tables or graphs.

For the purposes of reconciling the population sizes for a type (ii) experiment tabulated in the $P_g = .05$ column of Table 2 with those computed earlier for a type (i) experiment, the genotypic distribution for the best of k selected phenotypes was computed with $h = \frac{1}{2}$, $g = 1.645$, $x = 2$, and then compounded with the distribution of k for $N = 196$. Earlier computations had shown that when $h = \frac{1}{2}$ and $x = 2$ in a type (i) experiment, a population size of $N = 196$ is required for 90 percent confidence that at least one selection will belong to the top 5 percent of the genotypic distribution. Table 2, on the other hand, indicates that in a type (ii) experiment a population size of $N = 5300$ is required for 90 percent confidence that the phenotypically best selection will belong to the top 5 percent of the genotypic distribution.

Superficially, these two tabulated results might appear incompatible, however the conditions to be fulfilled are quite different. In the one case we require only that at least one of the selections be genetically superior, while in the other case we require that an identifiable one (the largest) of the selections be genetically superior. Clearly, the latter requirement is much more stringent and the population size necessary to achieve it is correspondingly much greater. For a population size of only 196 in a type (i) experiment with $h = \frac{1}{2}$ and $x = 2$ there is a 0.9 probability that at least one of the selections will have $g > 1.645$, but there is (by numerical integration) less than 0.65 probability that the phenotypically best selection will have $g > 1.645$.

In a type (ii) experiment, phenotypic ordering is an integral part of the selection process and thus contributes no additional information beyond that assumed in the planning stage. In this case the number (k) selected is fixed in advance, and the minimum selected phenotype x_k is a chance variable corresponding to the selection point x of a type (i)

experiment. In fact, if x_k were actually measured while the remaining top k-1 phenotypes were merely ranked then the information on these k-1 individuals would be of a type identical to that obtained on the k selections of a type (i) experiment, with x_k now playing the role of x.

Most commonly, all selected phenotypes will be measured, so that the outcome of the selection experiment consists of phenotypic observations $x_1 > x_2 > \cdots > x_k$, and Figures 1-3 then suffice to describe the probability distribution of genotypic values associated with each of these phenotypes, regardless of which type of experiment was employed.

Expected Identifiable Numbers of Genetic Deviates

Powers (1945), Powers et al., (1958) and Dudley and Powers (1959) introduced the concept of identifiable numbers of genetic deviates in the phenotypic classes of a segregating population. The identifiable numbers of genetic deviates are represented by the differences between these class frequencies and those of a nonsegregating population of the same size superimposed on the same population mean. As noted by Federer, Powers, and Payne (in process of publication) in the normal case the class frequencies of the segregating population exceed those of the corresponding nonsegregating population at a distance of

$$z = z' \sqrt{\sigma_g^2 + \sigma_e^2} = \sigma_e \sqrt{\frac{\log_e (1-h)}{-h}}$$

or more on either side of the mean. Consequently, in the sense defined by Powers, the frequency of identifiable genetic deviates is positive in any phenotypic class beginning at least a distance z away from the mean of an infinite population. In particular, in the entire tail of the distribution from z to $+\infty$ the frequency of identifiable genetic deviates is

$$P^+ = \bar{\Phi}\left(\sqrt{\frac{\log_e (1-h)}{-h}}\right) - \bar{\Phi}\left(\sqrt{\frac{(1-h)\log_e (1-h)}{-h}}\right)$$

$$= \bar{\Phi}\left(\frac{z'}{\sqrt{1-h}}\right) - \bar{\Phi}(z')$$

so that in a segregating population of size N the expected number of superior identifiable genetic deviates is defined as NP^+.

Unfortunately, in the present context this definition raises certain ambiguities, for in this same phenotypic interval (z, ∞) the proportion of genotypic values exceeding the mean is

$$P_{0 \cdot z'} = \frac{1}{1-\bar{\Phi}(z')} \int_{z'}^{\infty} \bar{\Phi}\left(\frac{y\sqrt{h}}{\sqrt{1-h}}\right) d\bar{\Phi}(y)$$

so that $N P_{0 \cdot z'} (1-\bar{\Phi}(z'))$ may also be described as the expected number of "superior" genotypes falling in this interval. Confusion may be avoided here by regarding P^+ as an index of heritability, comparable to the heritability ratio h, rather than attempting to interpret P^+ as a probability; in fact, for the normal case, P^+ is a monotone (increasing) function of h and is therefore equivalent to h as a heritability index. When the genotypic distribution is non-normal then, of course, the two indices P^+ and h are no longer equivalent and, in general, neither can be regarded as an adequate index of heritability in the sense of uniquely determining the genotypic distribution for a given phenotypic distribution. If a frequency difference is computed for every phenotypic class interval, however, all of the information in the genetic experiment is retained so that in this extreme if trivial form the P^+ index has optimum properties for any distribution model.

Numerical Illustrations With Sugarbeet Data

Experimental data fulfilling all of the requirements of the normal model are extremely uncommon, and in the strictest sense are actually non-existent. Aside from the problem of achieving normality and of achieving constant environmental variance by appropriate choices of scale of measurement, there is also the practical problem of conducting a completely randomized experiment with a single individual per plot as called for by the simple model considered here. The data used here for illustrative purposes, obtained from a population genetic study on sugarbeets conducted in 1960, cannot be rigorously shown to satisfy any of the requirements of the model, and so the application of the model provides only a crude guide to the true genetic character of the data. Until more general methods of

genetic analysis are developed, however, a crude guide is all that can be expected and may serve as a valuable tool in the analyses of such data.

For details of the design of the experiment, a description of the populations studied, and the adjustment of the frequency distributions to eliminate variation due to replications and populations see Powers et al., (in process of publication). With the exception of the predicted values the data in table 3 were taken from this article. They determined sucrose percentage of the sugarbeet roots, transformed the data to the logarithmic scale, and made the adjustments mentioned above. The resulting frequency distributions are thus freed of replication and population mean effects and hence are subject only to within plot sources of genetic and environmental variation. Intra plot correlations, positive before adjustment, are negative in the adjusted observations. Their effects are not taken into account in the following analysis. The frequency distribution (see Powers et al., 1958 and Powers et al., in process of publication) present a skew appearance indicating that the environmental distribution depicted by the nonsegregating entry may change shape with each genotype.

First the correspondence between the observed frequency difference beyond the points of intersection of the segregating and the nonsegregating frequency distributions are considered. The method of identifying these points of intersection is given by Powers et al., (1958). The predicted frequency differences, based on the observed within plot heritability ratio and the normal theory formula for NP^+ are shown in table 3 for each of the segregating populations. For example, with a heritability ratio of $h = .60569$ and $N = 450$ the predicted frequency difference to the right of

$$z' = \sqrt{\frac{.39432 \log_e (.39432)}{-.60569}} = .77836$$

is $$NP^+ = 450 \left[\bar{\alpha} \left(\frac{.77836}{\sqrt{.39432}} \right) - \bar{\alpha} (.77836) \right]$$

$$= 450 [.8924 - .7808] = 50$$

Examination of table 3 reveals that this prediction corresponds fairly closely with observation for most entries, though a few major discrepancies reduce the correlation between observed and predicted to 0.7. A somewhat

Table 3 Comparison of observed and predicted identifiable numbers of genetic deviates. 1960 Group I, log percentage sucrose, N = 450. 1/

Population and entry	Heritability ratio h	Identifiable numbers of genetic deviates			
		Superior		Total	
		observed	predicted	observed	predicted
CMS X 4W-34, 1	.60569	45	50	87	100
CMS X A54-1, 2	.62211	52	52	97	104
A54-1, 3	.63228	60	53	108	106
CMS X 4W-34 S_2, 4	.48490	42	35	75	70
4W-34 S_2, 5	.59884	45	49	77	97
CMS X 4W-34 asexual, recurrent, 6	.52778	44	40	82	81
4W-34 asexual, recurrent, 7	.41712	25	29	52	58
52-430 X 54-520 F_1, 8	.41019	32	29	64	57
54-520 X 52-305 F_1, 10	.35738	33	24	64	48
A56-3, 11	.51308	56	39	94	78
54-520, 12	.67032	44	59	83	118
Total		478	459	883	917
SS.		21864	20499	73541	81847
SCP		20794		76459	
C.T.		20771.27	19152.82	70880.82	76444.45
		19945.64		73610.09	
		1092.73	1346.18	2660.18	5402.55
		848.36		2848.91	
		r = .6995		r = .7515	

1/ With exception of the predicted values the data are taken from Powers, Remmenga, and Urquhart (in process of publication).

surprising side result is the high correlation of .99976 between h and NP^* in this table, indicating that despite this analytically complicated form, NP^* is a nearly linear function of h. In light of these many clear and suspected violations of the model, the degree of fit to the model predictions are somewhat surprising, and lend some credence to those assumptions which cannot be directly checked.

Another direct measure of departure from normality is given in table 4, where the observed frequencies in the upper tails of the distributions $(X > x')$ are compared to predicted frequencies $(N(1-\Phi(x')))$. Here the skewness of the phenotypic distributions is made more apparent, the predicted upper tail frequencies almost always exceeding the observed. The a posteriori probability of a superior genetic deviate in the tail, computed by interpolation from figures 4-12, would therefore appear to overestimate the true probability and hence have been adjusted downward. For example, in the ragion $X > \bar{g} + 1.2667\sqrt{\sigma_g^2 + \sigma_e^2}$ of a normal segregating population with h = .60569 the proportion of genotypes falling in the first quartile of the (normal) genotypic distribution is given by linear interpolation between the points $P_g = 25$, $x = 1.2667$ in figures 9 and 10. For h = .6 in figure 9 at $x = 1.2667$ and $P_g = 25$ we find $P_{g.x} = .825$ and for h = .7 in figure 10, $P_{g.x} = .887$. Interpolation to h = .60569 then gives

$$P_{g.x} = .60569\ (.887 - .825) + .825 = .828$$

and since 32 individuals instead of the predicted 46 belonged to this tail of the observed (standardized) phenotypic distribution then the predicted number of genotypes

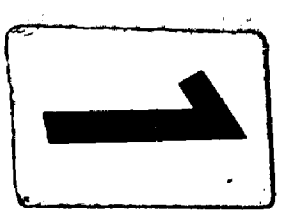
$$32\ P_{g.x} = 32\ (.828) = 26$$

was arbitrarily adjusted downward by the factor 32/46, giving

$$26\left(\frac{32}{46}\right) = 18$$

Table 4 Normal theory predicted proportions of superior genetic deviates based on adjustment 3, combining Group I, N = 450, D. F. = 420.

Population and entry	Heritability ratio	Standardized lower limit	Normal tail area	Normal tail number ± 2σ	Observed tail number	Predicted genotypes from the top half of the genotypic distribution		
	h^2		$p = 1-\Phi(x')$	$N_p \pm 2\sqrt{N_p(1-p)}$		proportion	number	adjusted
CMS X 4W-34, 1	0.60568649	.6000	.2742	123 ± 19	112	.902	101	92
CMS X A54-1, 2	0.62210681	.5875	.2785	125 ± 19	119	.908	108	103
A54-1, 3	0.63228438	.5794	.2812	127 ± 19	127	.910	116	116
CMS X 4W-34 S_2, 4	0.48489888	.6858	.2464	111 ± 18	109	.870	95	93
4W-34 S_2, 5	0.59884238	.6052	.2725	123 ± 19	112	.902	101	92
CMS X 4W-34 asexual, recurrent, 6	0.52778156	.6566	.2557	115 ± 18	111	.883	98	95
4W-34 asexual, recurrent, 7	0.41712484	.7295	.2328	105 ± 18	92	.848	78	68
52-430 X 54-520 F_1, 8	0.41018658	.7339	.2317	104 ± 18	99	.846	84	80
52-430 X 52-307 F_1, 9		.9555	.1697	76 ± 16	67			
54-520 X 52-305 F_1, 10	0.35737880	.7660	.2218	99 ± 18	101	.326	83	85
A56-3, 11	0.51308059	.6668	.2524	114 ± 18	123	.879	108	117
54-520, 12	0.67031556	.5487	.2917	131 ± 19	111	.920	102	86



ability and population genetic studies 1960, percentage sucrose transformed to logarithms.

Standardized lower limit	Normal tail area	tail number ± 2σ	Observed tail number	Predicted genotypes from the top quarter of the genotypic distribution		
x'	$p = 1-\Phi(x')$	$N_p \pm 2/\sqrt{N_p(1-p)}$	----	proportion	number	adjusted
1.2667	.1026	46 ± 13	32	.828	26	18
1.2402	.1075	48 ± 13	30	.822	25	16
1.2232	.1106	50 ± 13	32	.834	27	17
1.4478	.0738	33 ± 11	23	.793	18	13
1.2776	.1007	45 ± 13	34	.827	28	21
1.3861	.0829	37 ± 12	31	.807	25	21
1.5401	.0618	28 ± 10	20	.766	15	11
1.5493	.0607	27 ± 10	19	.763	14	10
2.0172	.0218	10 ± 6	7			
1.6171	.0529	24 ± 9	19	.735	14	11
1.4077	.0796	36 ± 11	24	.803	19	13
1.1583	.1234	56 ± 14	35	.841	29	18

Literature Cited

(1) Dudley, J. W. and Powers, LeRoy. 1959. Population genetic studies on sodium and potassium in sugar beets (Beta vulgaris L.). Jour. Amer. Soc. Sugar Beet Tech. XI(2):97-127.

(2) Eisenhart, C. 1947. The assumptions underlying the analysis of variance. Biometrics 3(1):1-21.

(3) Federer, W. T., Powers, LeRoy, and Payne, Merle G., (In process of publication) Studies on statistical procedures applied to chemical genetic data from sugar beets.

(4) Powers, LeRoy. 1945. Strawberry breeding studies involving crosses between the cultivated varieties (X Fragaria ananassa) and the native Rocky Mountain strawberry (F. ovalis). Jour. Agr. Res. 70:95-122.

(5) Powers, LeRoy, Robertson, D. W., and Clark, A. G. 1958. Estimation by the partitioning method of the numbers and proportions of genetic deviates in certain classes of frequency distributions. Jour. Amer. Soc. Sugar Beet Tech. IX(8):677-696.

(6) Powers, LeRoy, Remmenga, E. E., and Urquhart, N. S. (In process of publication). The partitioning method of genetic analysis applied to a study of weight per root and percentage sucrose in sugarbeets (Beta vulgaris L.).

Appendix

If $X = G + E$ where G and E are independent normal chance variables with means $\bar{g}$ and 0 and variances σ_g^2 and σ_e^2, respectively, then the conditional distribution of G for fixed X is normal with

$$\text{ave } (G|x) = \text{ave } (G) + \frac{\text{cov}(X,G)}{\text{var } (X)} \quad [x - \text{ave } (x)]$$

$$= \bar{g} + \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} \quad (x - \bar{g})$$

$$\text{var } (G|x) = \text{var } (G) \left[1 - \frac{[\text{cov}(X,G)]^2}{\text{var } (X) \text{ var } (G)}\right]$$

$$= \sigma_g^2 \left(1 - \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}\right)$$

Consequently, for

$$h = \frac{\sigma_e^2}{\sigma_e^2 + \sigma_g^2} \qquad x' = \frac{x - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}} \qquad g' = \frac{g - \bar{g}}{\sigma_g}$$

the conditional distribution of G is expressible as

$$P(G < g|x) = P\left[\frac{G - \text{ave}(G|x)}{\sqrt{\text{var}(G|x)}} < \frac{g - \text{ave } (G|x)}{\sqrt{\text{var}(G|x)}}\right] \tag{1}$$

$$= \bar{\Phi}\left[\frac{g' - \sqrt{h}x'}{\sqrt{(1-h)}}\right] = 1 - \bar{\Phi}\left[\frac{\sqrt{h}x' - g'}{\sqrt{(1-h)}}\right]$$

where $\bar{\Phi}$ is the standard (cumulative) normal distribution function.

Similarly, under the condition $X > x$

$$1 - P_{g \cdot x} = P(G < g|X > x) = \frac{1}{P(X > x)} \int_x^\infty P(G < g|y)\, dP(X < y)$$

$$= \frac{1}{1 - \bar{\Phi}(x')} \int_{x'}^\infty \bar{\Phi}\left(\frac{g' - y\sqrt{h}}{\sqrt{(1-h)}}\right) d\,\bar{\Phi}(y) \tag{2}$$

For the special case $g = \bar{g}$ and $h = \sqrt{h(1-h)} = \frac{1}{2}$, this expression reduces to

$$P(G < \bar{g} | X > x) = \frac{1}{1-\bar{\Phi}(x')} \int_{x'}^{\infty} \bar{\Phi}(-y)\, d\bar{\Phi}(y)$$

$$= 1 - \frac{1}{1-\bar{\Phi}(x')} \int_{x'}^{\infty} \bar{\Phi}(y)\, d\,\bar{\Phi}(y)$$

$$= 1 - \tfrac{1}{2}\,[1 + \bar{\Phi}(x')]$$

while (1) becomes simply

$$P(G < \bar{g} | x) = 1 - \bar{\Phi}(x')$$

If $X_N < \ldots < X_1$ denote the ranked observations in a random sample of size N then the distribution of the G- component of X_k is given by

$$P(G_k < g) = \int_{-\infty}^{\infty} P(G < g | x)\, dP(X_k < x)$$

$$= k \int_{-\infty}^{\infty} \bar{\Phi}\left(\frac{g' - x'\sqrt{h}}{\sqrt{(1-h)}}\right) \binom{N}{k} [\bar{\Phi}(x')]^{N-k} [1-\bar{\Phi}(x')]^{k-1} d\bar{\Phi}(x')$$

For the special case $g = \bar{g}$ and $h = \frac{1}{2}$ this gives

$$P(G_k < \bar{g}) = k \int_{-\infty}^{\infty} [1 - \bar{\Phi}(x')] \binom{N}{k} [\bar{\Phi}(x')]^{N-k} [1 - \bar{\Phi}(x')]^{k-1} d\bar{\Phi}(x')$$

$$= \frac{k}{N+1}$$

so the expected number of G- components less than $\bar{g}$ among the k largest observations $X_k < \ldots < X_1$ is the sum

$$\sum_{j=1}^{k} \frac{j}{N+1} = \frac{k\,(k+1)}{2\,(N+1)}$$

Let

$$T_i = T(G_i, X_i; g, x) = \begin{cases} 1 \text{ if } \dfrac{X_i - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_s^2}} > x \text{ and } \dfrac{G_i - \bar{g}}{\sigma_g} > g \\ 0 \text{ otherwise} \end{cases}$$

then

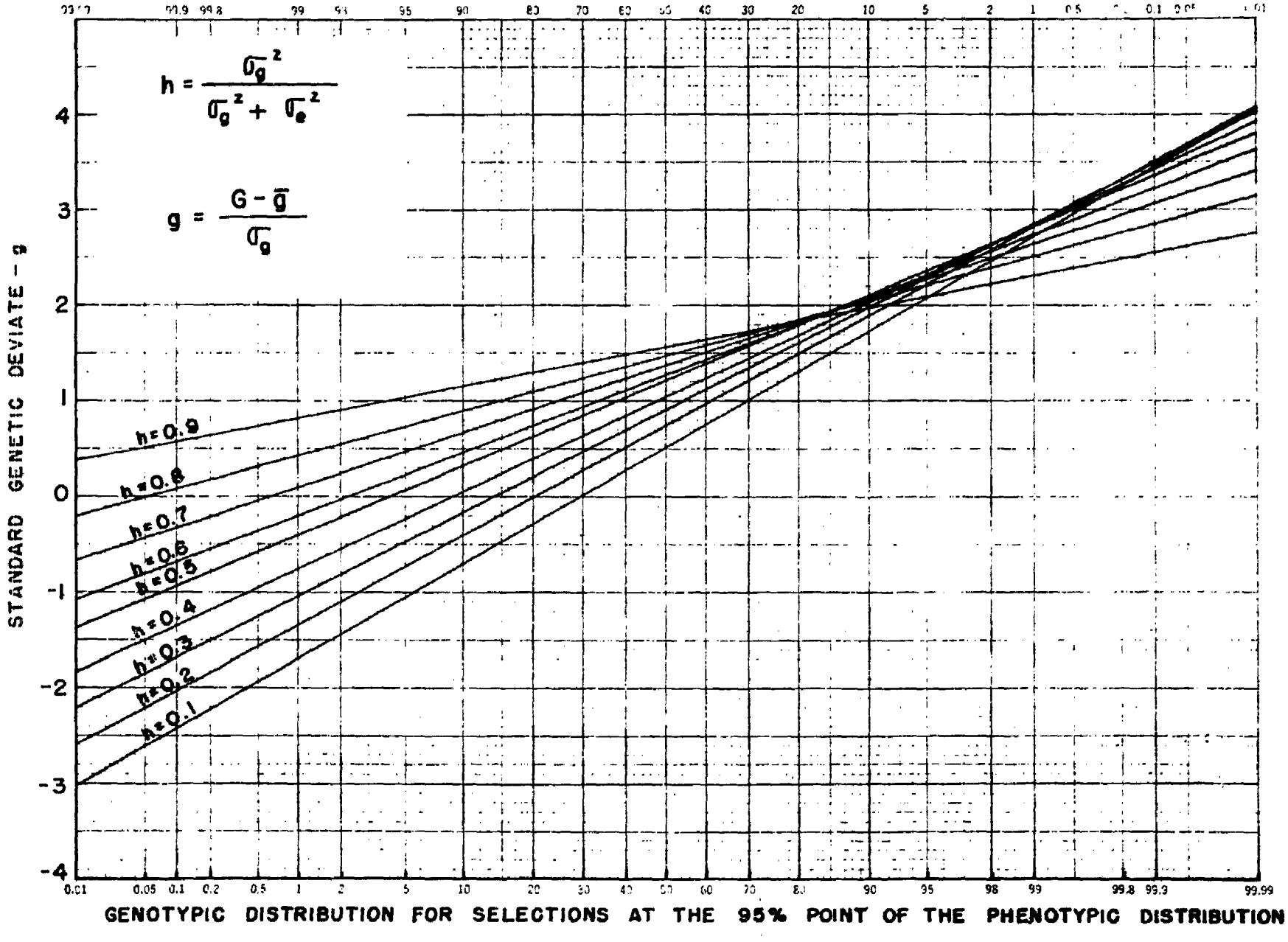
$$P(T_i = 1) = \int_x^\infty \bar{\Phi}\left(\frac{y\sqrt{h} - g}{\sqrt{(1-h)}}\right) d\bar{\Phi}(y) \equiv P_{g,x}$$

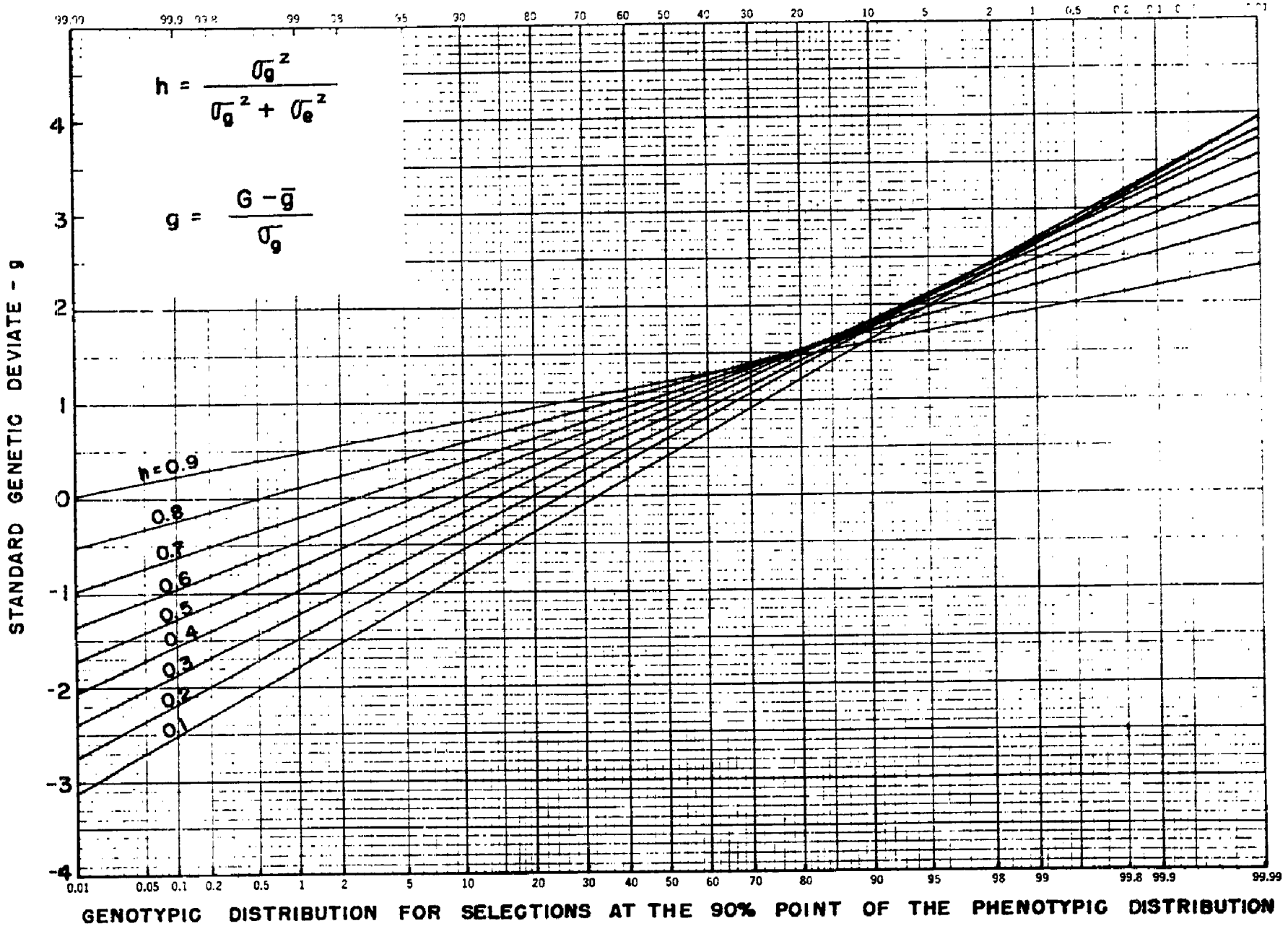
and

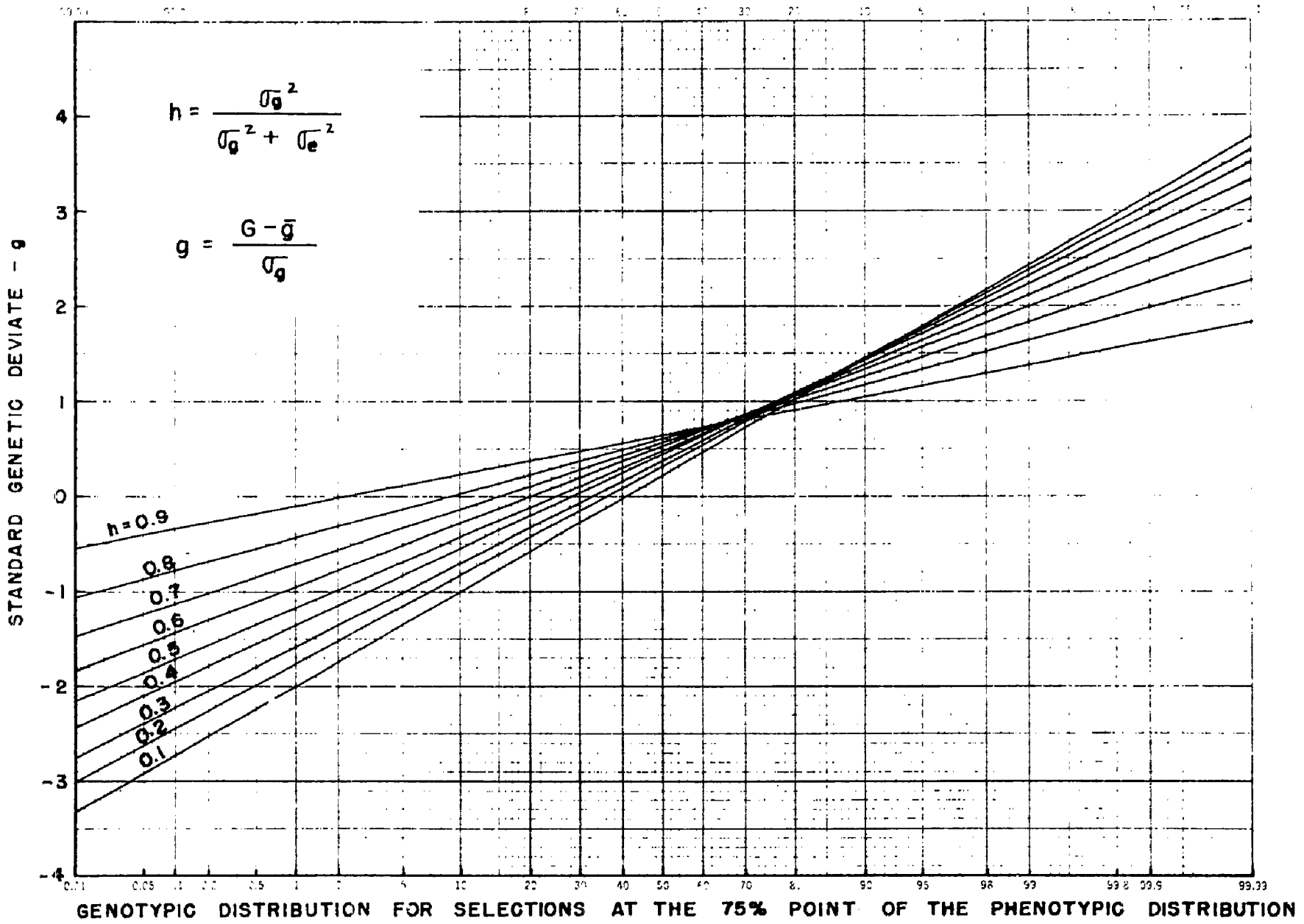
$$P\left(\sum_1^N T_i = t\right) = \binom{N}{t} P_{g,x}^t (1-P_{g,x})^{N-t}.$$

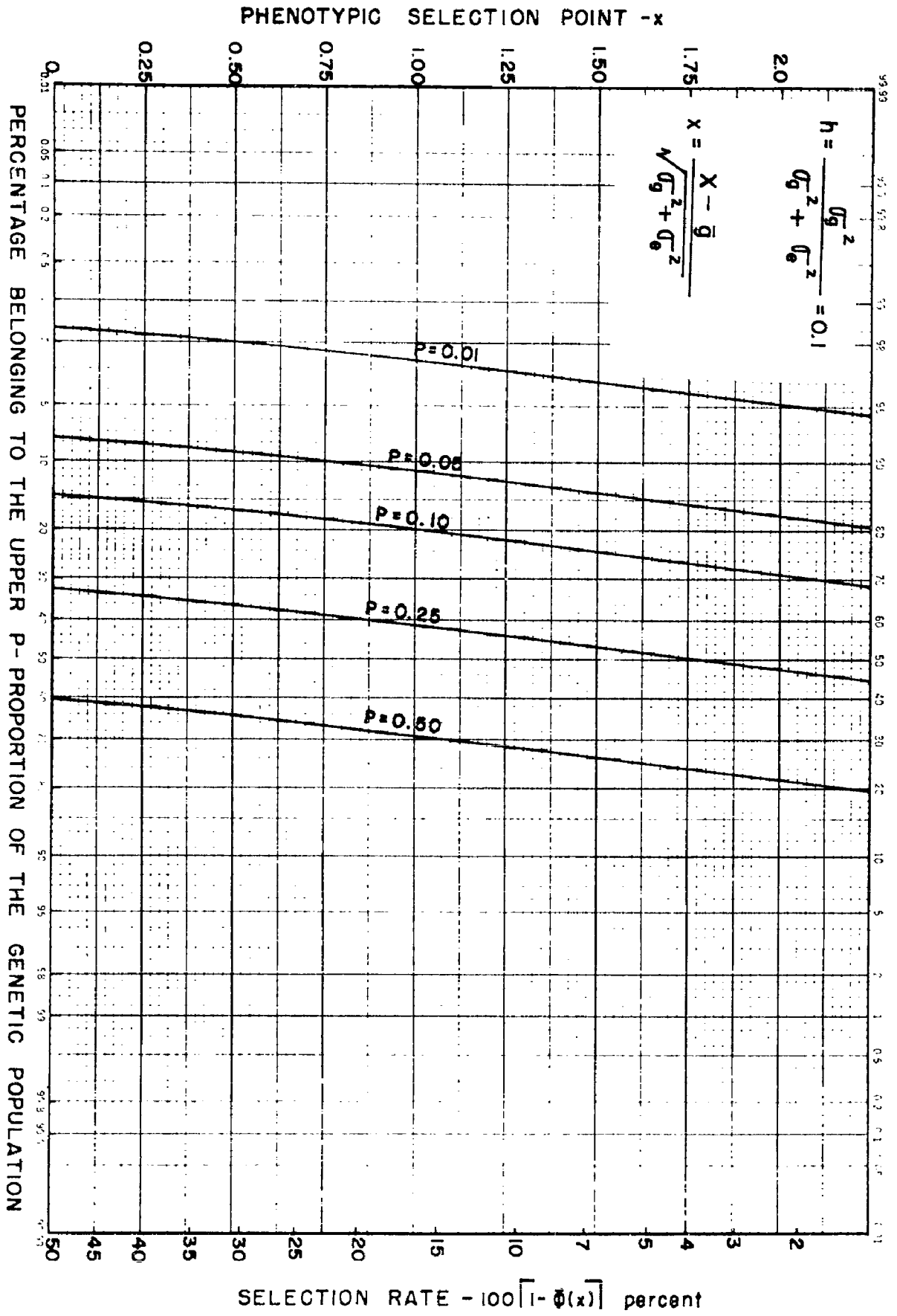
In particular,

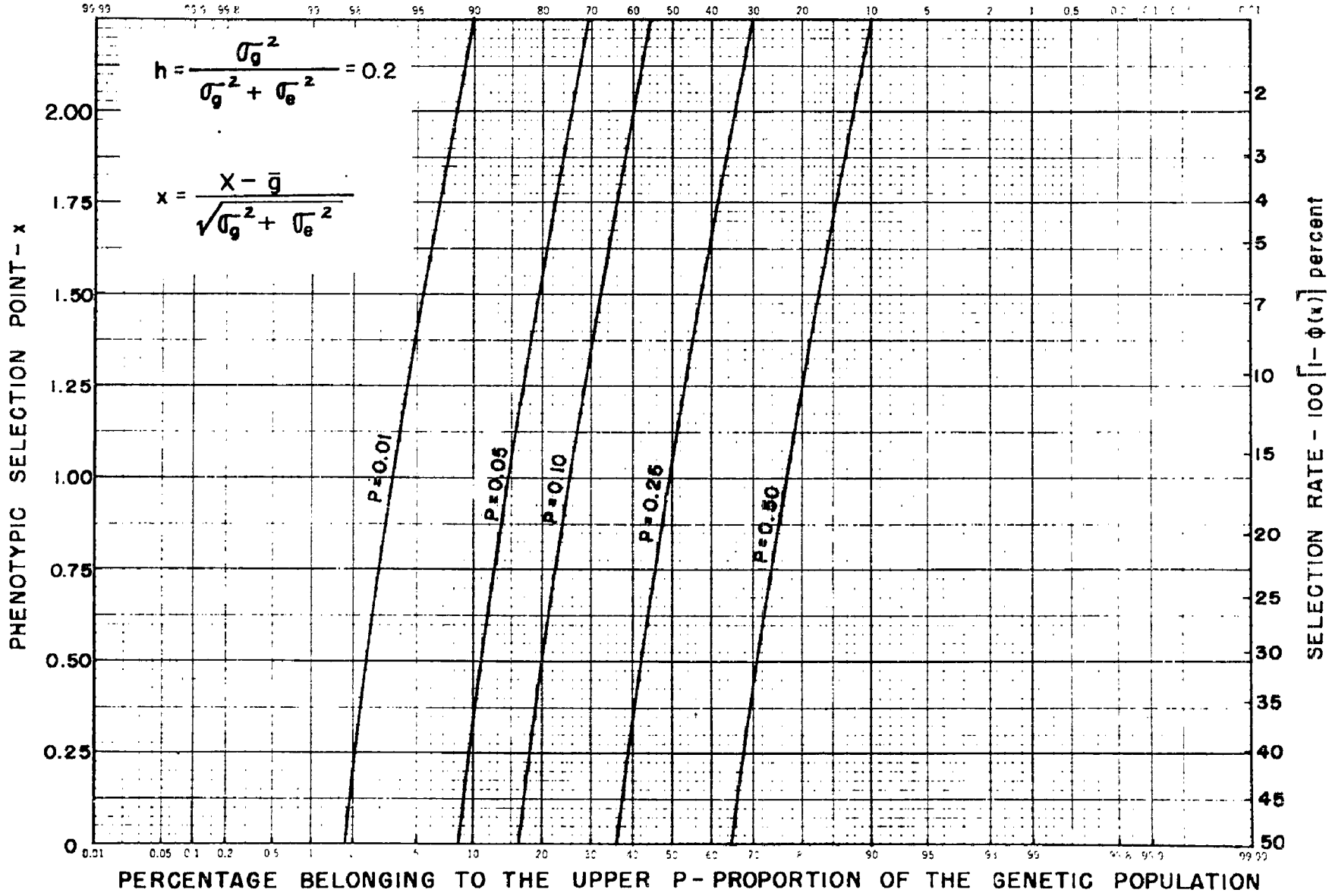
$$P\left(\sum_1^N T_i \geq 1\right) = 1 - (1-P_{g,x})^N$$

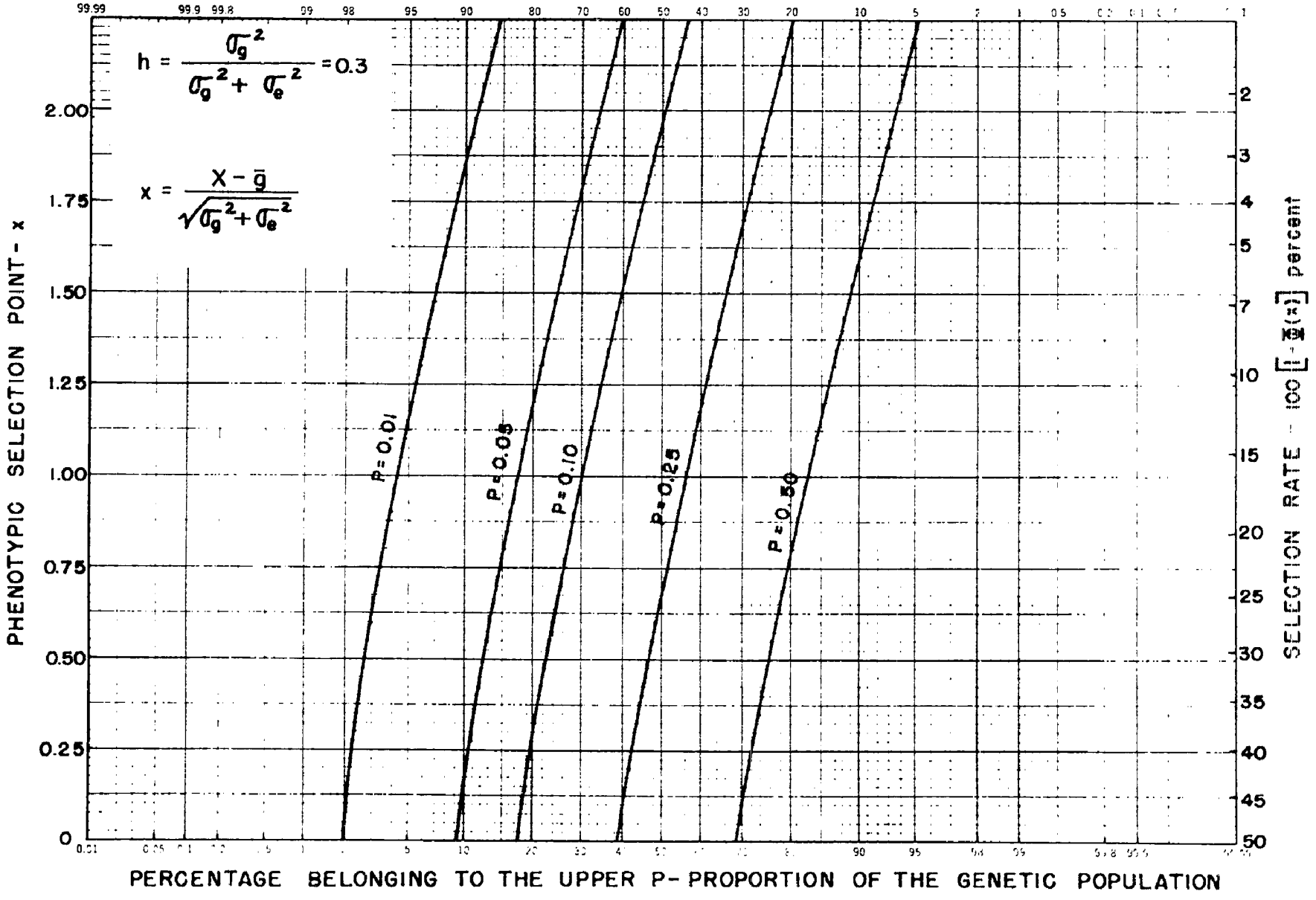

$h = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$
$g = \frac{G - \bar{g}}{\sigma_g}$
h=0.9
h=0.8
h=0.7
h=0.6
h=0.5
h=0.4
h=0.3
h=0.2
h=0.1
STANDARD GENETIC DEVIATE - g
4
3
2
1
0
-1
-2
-3
-4
0.01 0.05 0.1 0.2 0.5 1 2 5 10 20 30 40 50 60 70 80 90 95 98 99 99.8 99.9 99.99
GENOTYPIC DISTRIBUTION FOR SELECTIONS AT THE 95% POINT OF THE PHENOTYPIC DISTRIBUTION

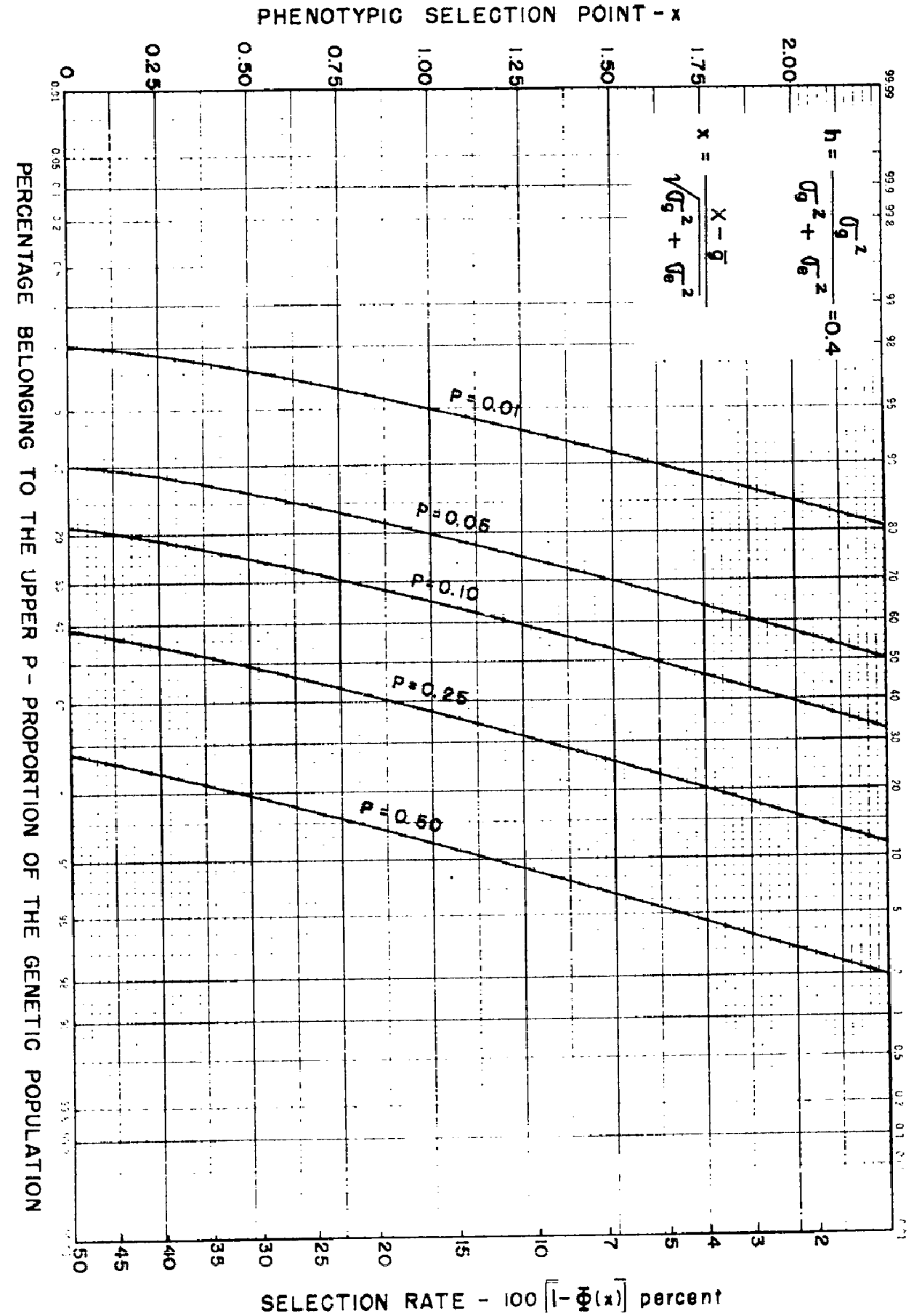

$h = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$
$g = \frac{G - \bar{g}}{\sigma_g}$
h=0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
STANDARD GENETIC DEVIATE - g
4
3
2
1
0
-1
-2
-3
-4
99.99 99.9 99.8 99 98 95 90 80 70 60 50 40 30 20 10 5 2 1 0.5
0.01 0.05 0.1 0.2 0.5 1 2 5 10 20 30 40 50 60 70 80 90 95 98 99 99.8 99.9 99.99
GENOTYPIC DISTRIBUTION FOR SELECTIONS AT THE 90% POINT OF THE PHENOTYPIC DISTRIBUTION

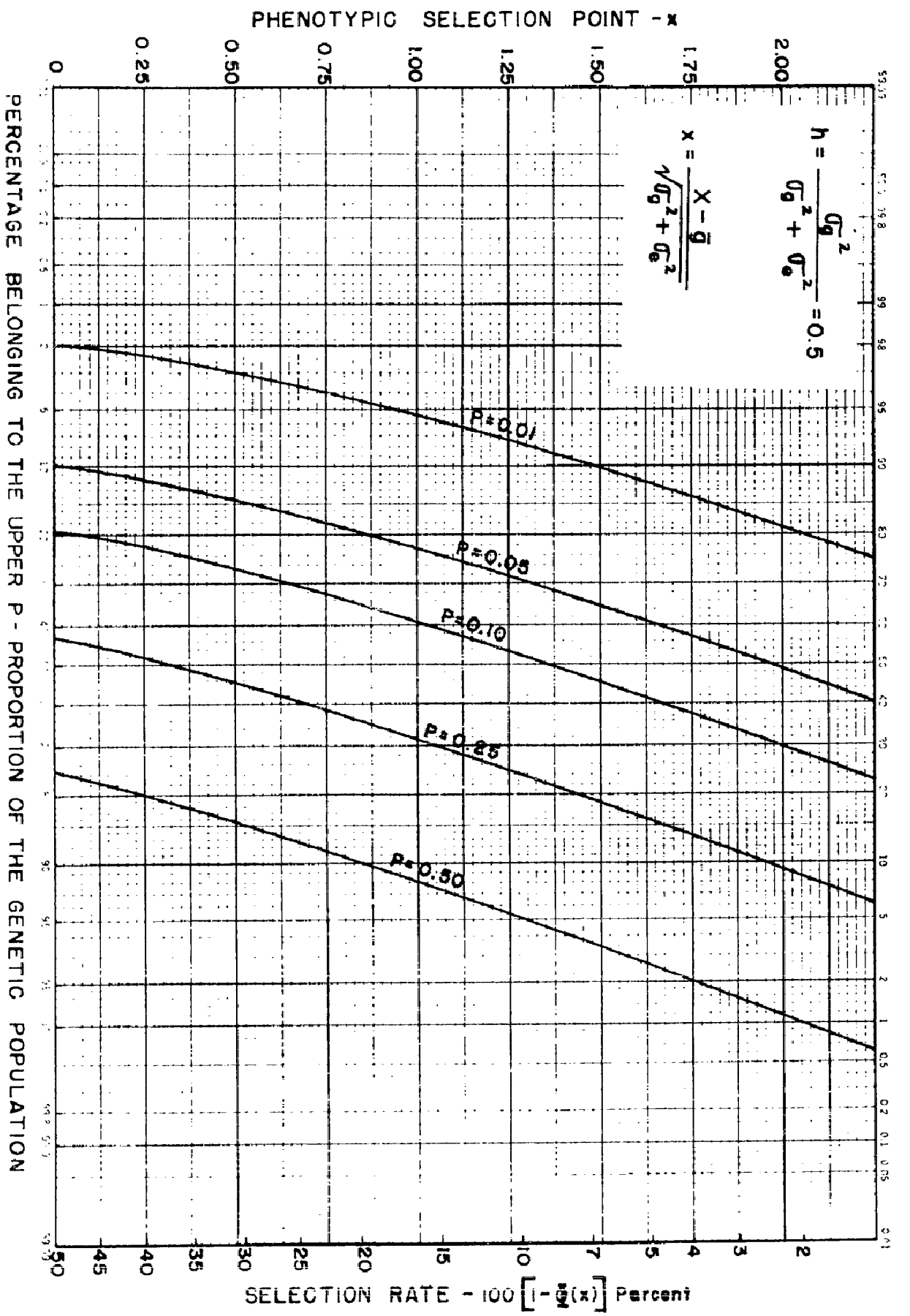

$h = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$
$g = \frac{G - \bar{g}}{\sigma_g}$
h=0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
STANDARD GENETIC DEVIATE - g
4
3
2
1
0
-1
-2
-3
-4
0.01
0.05
99.33
GENOTYPIC DISTRIBUTION FOR SELECTIONS AT THE 75% POINT OF THE PHENOTYPIC DISTRIBUTION

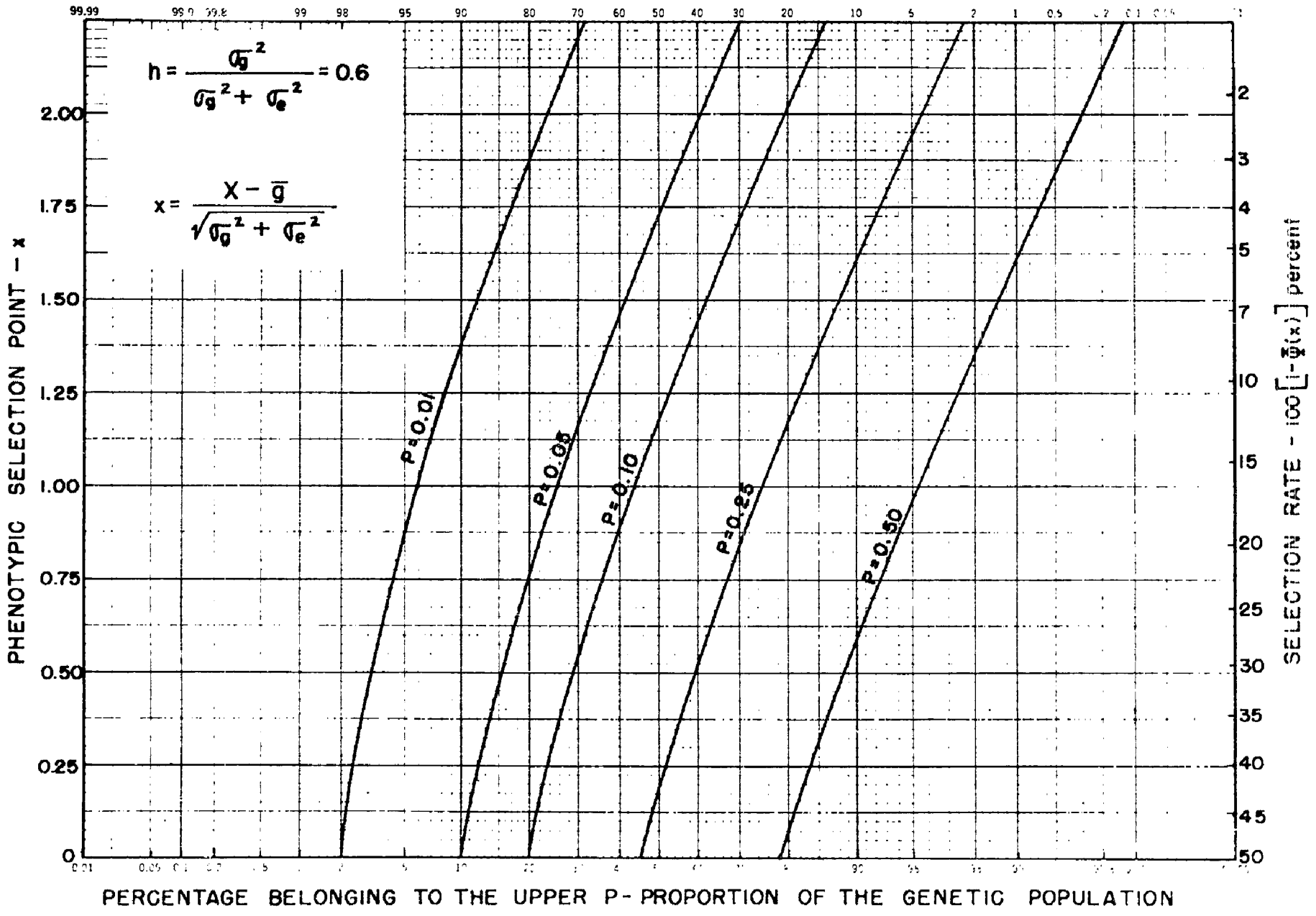

PHENOTYPIC SELECTION POINT -x
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.0
$h = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} = 0.1$
$x = \frac{X - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}}$
P=0.01
P=0.05
P=0.10
P=0.25
P=0.50
PERCENTAGE BELONGING TO THE UPPER P- PROPORTION OF THE GENETIC POPULATION
SELECTION RATE - $100[1-\Phi(x)]$ percent

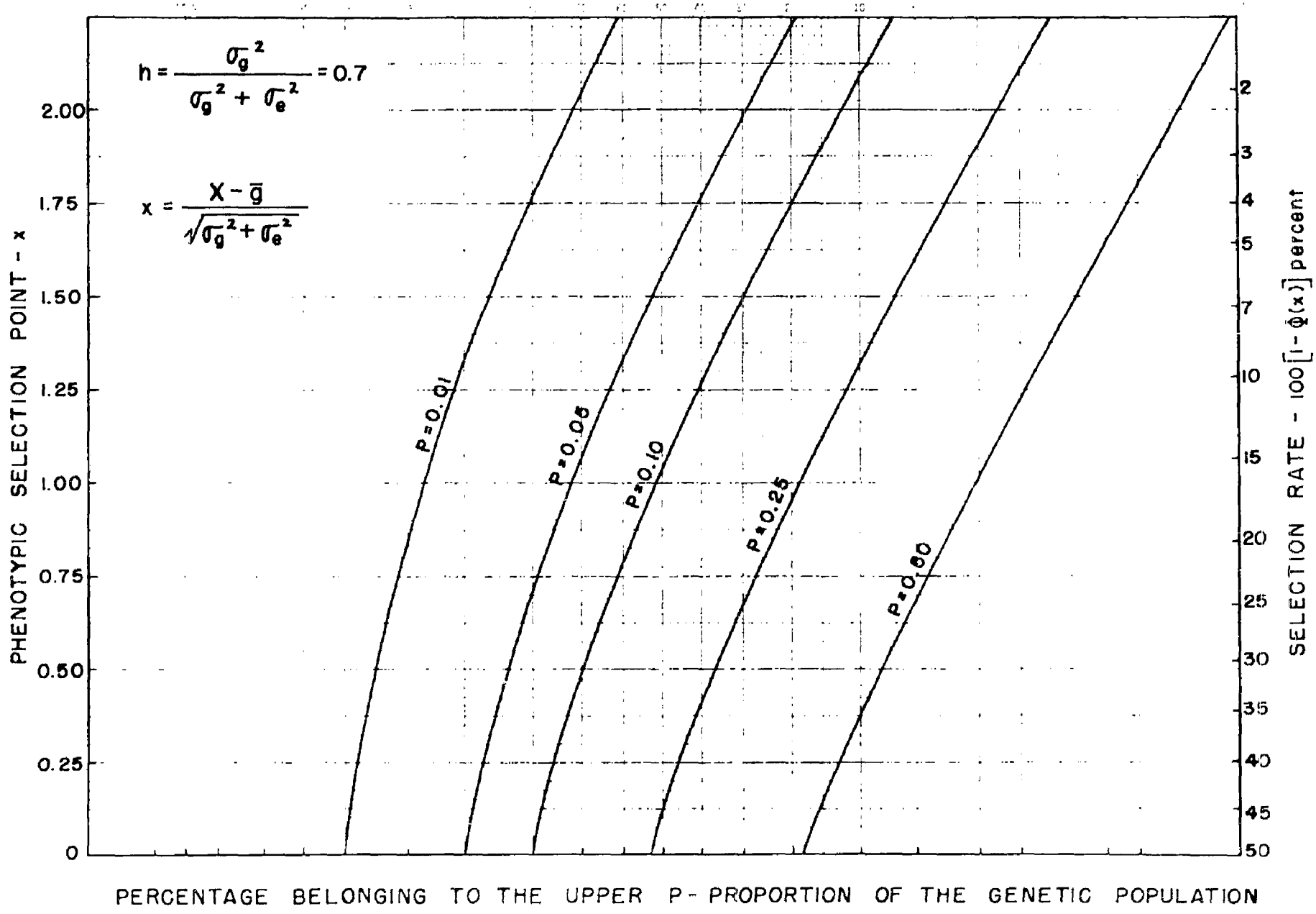

$h = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} = 0.2$
$x = \frac{X - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}}$
P=0.01
P=0.05
P=0.10
P=0.25
P=0.50
PHENOTYPIC SELECTION POINT - x
SELECTION RATE - 100[1 - φ(x)] percent
PERCENTAGE BELONGING TO THE UPPER P - PROPORTION OF THE GENETIC POPULATION

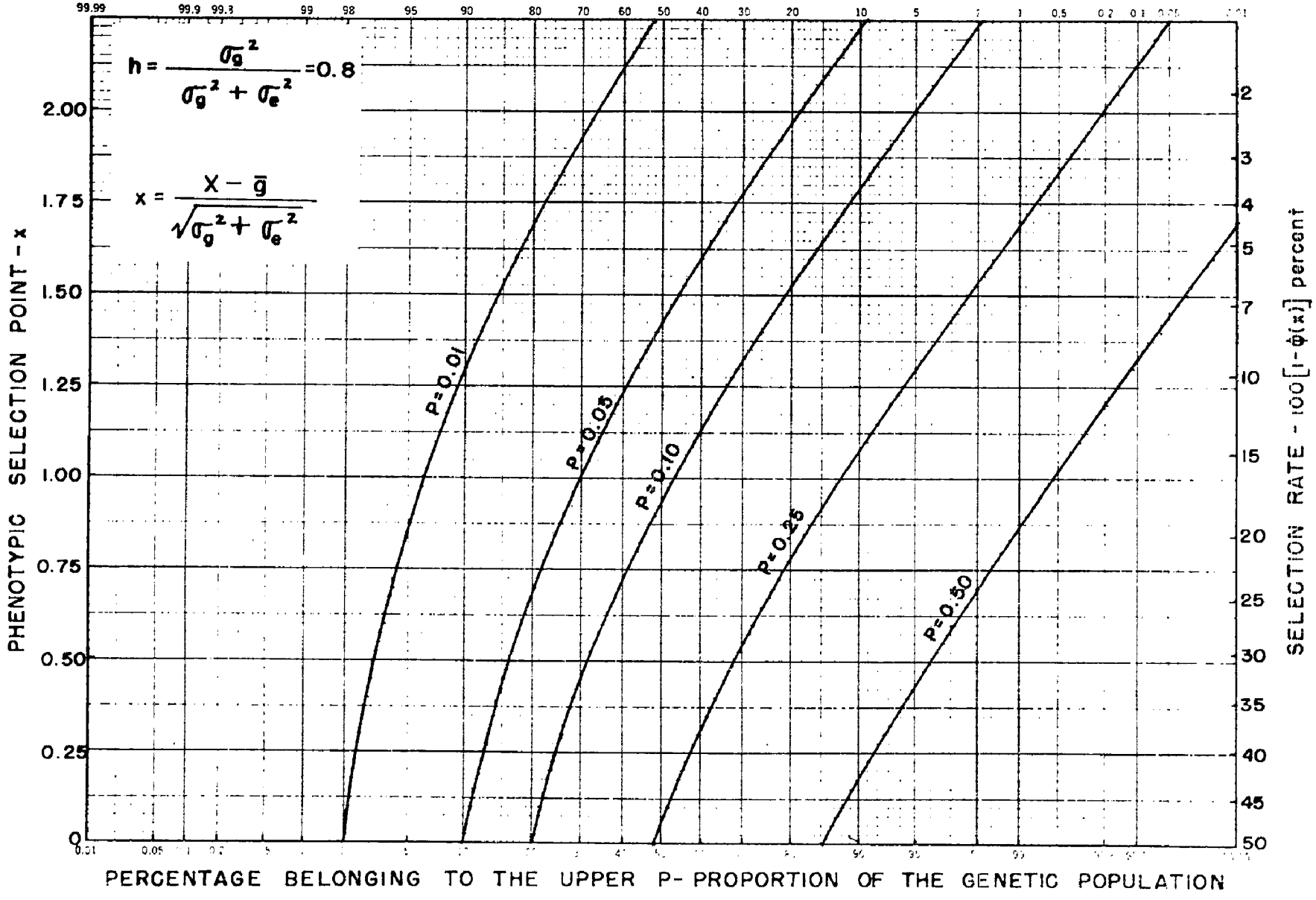

$h = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} = 0.3$
$x = \frac{X - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}}$
P = 0.01
P = 0.05
P = 0.10
P = 0.25
P = 0.50
PHENOTYPIC SELECTION POINT - x
SELECTION RATE - 100 [1 - Φ(x)] percent
PERCENTAGE BELONGING TO THE UPPER P - PROPORTION OF THE GENETIC POPULATION

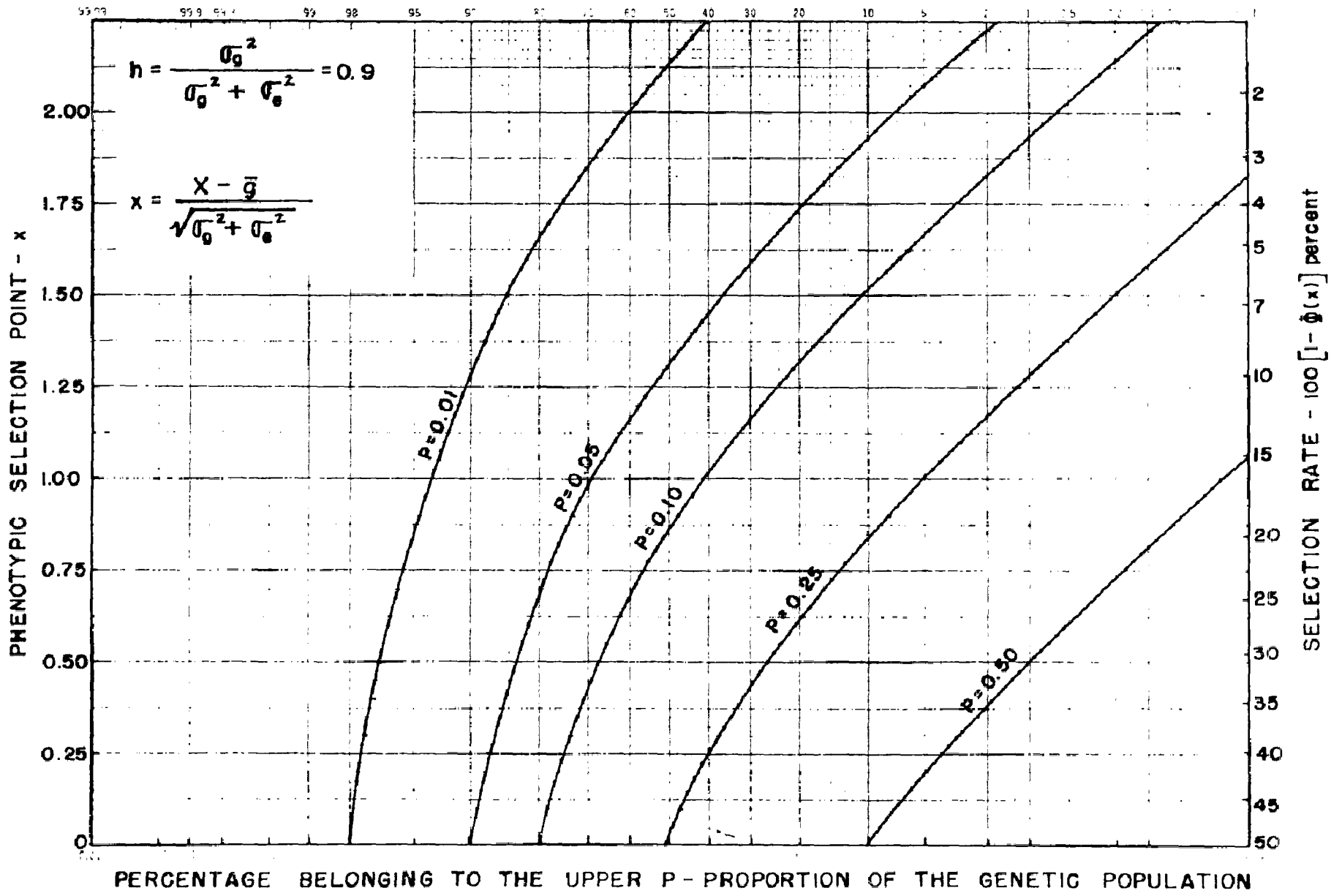

PHENOTYPIC SELECTION POINT - x
0
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
$h = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} = 0.4$
$x = \frac{X - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}}$
P = 0.01
P = 0.05
P = 0.10
P = 0.25
P = 0.50
PERCENTAGE BELONGING TO THE UPPER P - PROPORTION OF THE GENETIC POPULATION
SELECTION RATE - $100\left[1 - \Phi(x)\right]$ percent

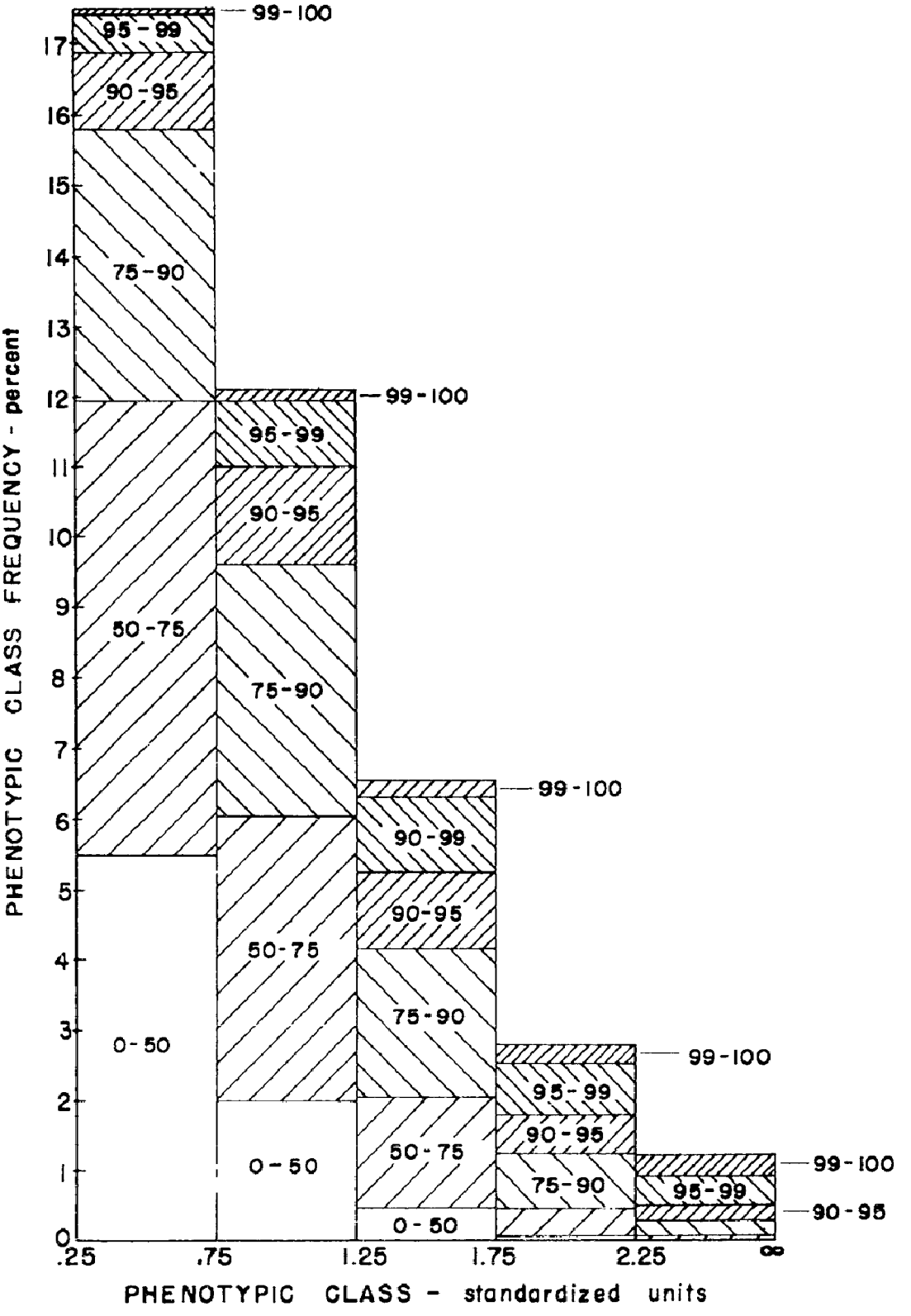

PHENOTYPIC SELECTION POINT - x
0
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
$h = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} = 0.5$
$x = \frac{X - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}}$
P=0.01
P=0.05
P=0.10
P=0.25
P=0.50
50
45
40
35
30
25
20
15
10
7
5
4
3
2
SELECTION RATE - 100 [1 - Φ(x)] Percent
PERCENTAGE BELONGING TO THE UPPER P - PROPORTION OF THE GENETIC POPULATION


$h = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} = 0.6$
$x = \frac{X - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}}$
P=0.01
P=0.05
P=0.10
P=0.25
P=0.50
PHENOTYPIC SELECTION POINT - x
SELECTION RATE - 100 [1-Φ(x)] percent
PERCENTAGE BELONGING TO THE UPPER P-PROPORTION OF THE GENETIC POPULATION


$h = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} = 0.7$
$x = \frac{X - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}}$
PHENOTYPIC SELECTION POINT - x
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0
P=0.01
P=0.05
P=0.10
P=0.25
P=0.50
2
3
4
5
7
10
15
20
25
30
35
40
45
50
SELECTION RATE - $100[1 - \Phi(x)]$ percent
PERCENTAGE BELONGING TO THE UPPER P - PROPORTION OF THE GENETIC POPULATION


$h = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} = 0.8$
$x = \frac{X - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}}$
P=0.01
P=0.05
P=0.10
P=0.25
P=0.50
PHENOTYPIC SELECTION POINT - x
SELECTION RATE - 100[1-Φ(x)] percent
PERCENTAGE BELONGING TO THE UPPER P- PROPORTION OF THE GENETIC POPULATION


$h = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} = 0.9$
$x = \frac{X - \bar{g}}{\sqrt{\sigma_g^2 + \sigma_e^2}}$
PHENOTYPIC SELECTION POINT - x
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0
P=0.01
P=0.05
P=0.10
P=0.25
P=0.50
SELECTION RATE - 100 [1 - Φ(x)] percent
2
3
4
5
7
10
15
20
25
30
35
40
45
50
PERCENTAGE BELONGING TO THE UPPER P - PROPORTION OF THE GENETIC POPULATION


99-100
95-99
90-95
75-90
50-75
0-50
PHENOTYPIC CLASS FREQUENCY - percent
PHENOTYPIC CLASS - standardized units
.25
.75
1.25
1.75
2.25
∞

CORNELL UNIVERSITY - BIOMETRICS UNIT

Distribution List for Unclassified Technical Reports

Contract Nonr-401(39)
Project (NR 042-212)

Head, Logistics and Mathematical Statistics Branch
Office of Naval Research
Washington 25, D. C. 3

Commanding Officer
Office of Naval Research
346 Broadway
New York 13, New York 1

ASTIA Document Service Center
Arlington Hall Station
Arlington 12, Virginia 10

Institute for Defense Analyses
Communications Research Division
von Neumann Hall
Princeton, New Jersey 1

Technical Information Officer
Naval Research Laboratory
Washington 25, D. C. 6

Professor T. W. Anderson
Department of Mathematical Statistics
Columbia University
New York 27, New York 1

Professor Z. W. Birnbaum
Laboratory of Statistical Research
Department of Mathematics
University of Washington
Seattle 5, Washington 1

Professor A. W. Bowker
Applied Mathematics and Statistics Laboratory
Stanford University
Stanford, California 1

Professor Ralph A. Bradley
Department of Statistics
Florida State University
Tallahassee, Florida 1

Dr. John W. Cell
Department of Mathematics
North Carolina State College
Raleigh, North Carolina 1

Professor Herman Chernoff
Applied Mathematics and Statistics Laboratory
Stanford University
Stanford, California 1

Professor W. G. Cochran
Department of Statistics
Harvard University
Cambridge, Massachusetts 1

Dr. C. Clark Cockerham
Institute of Statistics
North Carolina State College
Raleigh, North Carolina 1

Professor Cyrus Derman
Dept. of Industrial Engineering
Columbia University
New York 27, New York 1

Professor Benjamin Epstein
Applied Mathematics and Statistics Laboratory
Stanford University
Stanford, California 1

Dr. W. T. Federer
Biometrics Unit
Plant Breeding Department
Cornell University
Ithaca, New York 1

Dr. R. J. Freund
Department of Statistics and Statistical Laboratory
Virginia Polytechnic Institute
Blacksburg, Virginia 1

Professor H. P. Goode
Dept. of Industrial and Engineering Administration
Cornell University
Ithaca, New York 1

Professor W. Hirsch
Institute of Mathematical Sciences
New York 3, New York 1

Professor Harold Hotelling,
Associate Director
Institute of Statistics
University of North Carolina
Chapel Hill, North Carolina 1

Professor Oscar Kempthorne
Statistics Laboratory
Iowa State University
Ames, Iowa 1

Professor Gerald J. Lieberman
Applied Mathematics and
Statistics Laboratory
Stanford University
Stanford, California 1

Dr. Arthur E. Mace
Battelle Memorial Institute
505 King Avenue
Columb s 1, Ohio 1

Professor J. Neyman
Department of Statistics
University of California
Berkeley 4, California 1

Dr. F. Oberhettinger
Department of Mathematics
Oregon State College
Corvallis, Oregon 1

Professor Herbert Robbins
Mathematical Statistics Dept.
Fayerweather Hall
Columbia University
New York 27, New York 1

Professor Murray Rosenblatt
Department of Mathematics
Brown University
Providence 12, Rhode Island 1

Professor H. Rubin
Department of Statistics
Michigan State University
East Lansing, Michigan 1

Professor I. R. Savage
School of Business Administration
University of Minnesota
Minneapolis, Minnesota 1

Professor L. J. Savage
Statistical Research Laboratory
Chicago University
Chicago 37, Illinois 1

Professor W. L. Smith
Statistics Department
University of North Carolina
Chapel Hill, North Carolina 1

Professor Frank Spitzer
Department of Mathematics
University of Minnesota
Minneapolis, Minnesota 1

Dr. H. Teicher
Statistical Laboratory
Engineering Administration Building
Purdue University
Lafayette, Indiana 1

Professor M. B. Wilk
Statistics Center
Rutgers - The State University
New Brunswick, New Jersey 1

Professor S. S. Wilks
Department of Mathematics
Princeton University
Princeton, New Jersey 1

Professor J. Wolfowitz
Department of Mathematics
White Hall
Cornell University
Ithaca, New York 1